

An Inequality for the Correlation of Two Functions Operating on Symmetric Bivariate Normal Variables

Ran Hadad, Uri Erez, and Yaming Yu

Abstract

An inequality is derived for the correlation of two univariate functions operating on symmetric bivariate normal random variables. The inequality is a simple consequence of the Cauchy-Schwarz inequality.

I. INTRODUCTION

Statistical characterization of the output of non-linear systems operating on stochastic processes is in general difficult. Nonetheless, when the input process is Gaussian and the system is a memoryless non-linearity, several particularly simple and useful properties are known. Among these are Bussgang's theorem [1] and its generalizations (e.g., [2], [3]), results concerning the maximal correlation coefficient [4], as well as results on the output distortion-to-signal power ratio [5]. In the present note, we describe another simple result as described in the next section. A generalization to a more general class of random variables is described in Section III. An application is presented in Section IV.

II. STATEMENT OF RESULT AND PROOF

Lemma 1. *Let Z_1 and Z_2 be zero-mean bivariate normal random variables with variance σ^2 and correlation coefficient $\rho > 0$. Then,*

$$\mathbb{E}^2 [g_1(Z_1)g_2(Z_2)] \leq \mathbb{E} [g_1(Z_1)g_1(Z_2)] \mathbb{E} [g_2(Z_1)g_2(Z_2)],$$

R. Hadad and U. Erez with the Department of Electrical Engineering-Systems, Tel Aviv University, Ramat Aviv 69978, Israel; (e-mails: ranhadad@post.tau.ac.il, uri@eng.tau.ac.il).

Y. Yu is with the Department of Statistics, University of California, Irvine, CA 92697, USA; (e-mail: yamingy@uci.edu).

for any g_1 and g_2 for which the expectations exist, with equality if and only if g_1 and g_2 are equal up to a multiplicative constant.

Note that for $\rho = 1$ the lemma reduces to the standard probabilistic Cauchy-Schwarz inequality.

Proof. Consider the inner product of real functions defined by

$$\langle f, g \rangle \triangleq \mathbb{E}[f(Z)g(Z)]$$

where $Z \sim \mathcal{N}(0, \sigma^2)$.

According to Mehler's formula [6] (see also [7]) the joint density function of (Z_1, Z_2) may be written as,

$$\begin{aligned} f_{Z_1, Z_2}(z_1, z_2) &= \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} e^{-\frac{1}{2}\frac{z_1^2 - 2\rho z_1 z_2 + z_2^2}{\sigma^2(1-\rho^2)}} \\ &= \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2}\frac{z_1^2 + z_2^2}{\sigma^2}} \sum_{n=0}^{\infty} \frac{1}{n!} He_n\left(\frac{z_1}{\sigma}\right) He_n\left(\frac{z_2}{\sigma}\right) \rho^n, \end{aligned} \quad (1)$$

where $He_n(x)$ are the probabilists' Hermite polynomials defined as,

$$He_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{d^n}{dx^n} e^{-\frac{x^2}{2}}, \quad n \geq 0.$$

The Hermite polynomials constitute a complete orthogonal basis of polynomials with respect to the standard normal probability density function [8], so that

$$\langle \tilde{He}_n, \tilde{He}_m \rangle = \mathbb{E}[\tilde{He}_n(Z)\tilde{He}_m(Z)] = n! \delta_{n,m}$$

where $\tilde{He}_n(x) = He_n(\frac{x}{\sigma})$, and $\delta_{n,m}$ is the Kronecker delta function.

Let $a_{g_i, n} = \langle g_i, \tilde{He}_n \rangle$. The function $g_i(x)$ may be represented by the series

$$g_i(x) = \sum_{n=0}^{\infty} \frac{1}{n!} a_{g_i, n} He_n\left(\frac{x}{\sigma}\right), \quad i = 1, 2. \quad (2)$$

The following expectations are obtained by applying (1):

$$\begin{aligned} \mathbb{E}[g_1(Z_1)g_2(Z_2)] &= \sum_{n=0}^{\infty} \frac{1}{n!} a_{g_1, n} a_{g_2, n} \rho^n, \\ \mathbb{E}[g_i(Z_1)g_i(Z_2)] &= \sum_{n=0}^{\infty} \frac{1}{n!} a_{g_i, n}^2 \rho^n, \quad i = 1, 2. \end{aligned}$$

Using these identities and assuming $\rho > 0$ we have,

$$\begin{aligned}
\mathbb{E}^2 [g_1(Z_1)g_2(Z_2)] &= \left(\sum_{n=0}^{\infty} \frac{a_{g_1,n} a_{g_2,n}}{n!} \rho^n \right)^2 \\
&= \left(\sum_{n=0}^{\infty} \frac{a_{g_1,n} \rho^{\frac{n}{2}}}{\sqrt{n!}} \frac{a_{g_2,n} \rho^{\frac{n}{2}}}{\sqrt{n!}} \right)^2 \\
&\leq \left(\sum_{n=0}^{\infty} \frac{a_{g_1,n}^2 \rho^n}{n!} \right) \left(\sum_{n=0}^{\infty} \frac{a_{g_2,n}^2 \rho^n}{n!} \right) \\
&= \mathbb{E} [g_1(Z_1)g_1(Z_2)] \mathbb{E} [g_2(Z_1)g_2(Z_2)],
\end{aligned}$$

where the inequality follows by the CauchySchwarz inequality for sequences, which holds with equality only when $a_{g_1,n} = c \cdot a_{g_2,n}$ for some constant c , $\rho > 0$, and for all $n \geq 0$. Since the case of equality holds only when the coefficients in the series (2) are equal up to a multiplicative constant, it follows that equality holds only when g_1 and g_2 are equal up to a multiplicative constant. \square

When both functions are even or odd, we may apply the lemma to Z_1 and $-Z_2$ to obtain,

Corollary 1. *Let Z_1 and Z_2 be zero-mean bivariate normal random variables with variance σ^2 and correlation coefficient $\rho \neq 0$. Then, for g_1 and g_2 that are both even or odd functions,*

$$\mathbb{E}^2 [g_1(Z_1)g_2(Z_2)] \leq \mathbb{E} [g_1(Z_1)g_1(Z_2)] \mathbb{E} [g_2(Z_1)g_2(Z_2)],$$

for any such g_1 and g_2 for which the expectations exist, with equality if and only if g_1 and g_2 are equal up to a multiplicative constant.

Remark 1. *It is interesting to contrast the lemma with the maximal correlation property of normal vectors. Specifically, consider the case where Z_1 and Z_2 are zero mean and both g_1 and g_2 are odd, so that $\mathbb{E} [g_i(Z_j)] = 0$, for all $i = 1, 2$ and $j = 1, 2$. Then the maximal correlation property [4] yields the inequality*

$$\mathbb{E}^2 [g_1(Z_1)g_2(Z_2)] \leq \rho^2 \mathbb{E} [g_1^2(Z_1)] \mathbb{E} [g_2^2(Z_2)].$$

III. GENERALIZATION

The lemma may be generalized to a broader class of random variables (Z_1, Z_2) , where Z_1 and Z_2 result from passing some random variable Z through two independent realizations of the same “channel” (in information-theoretic terms). This generalization is stated in the next lemma. We note however that unlike Lemma 1, the “only if” condition for equality does not necessarily apply.

Lemma 2. *Let Z_1 and Z_2 be random variables such that for some random variable Z , Z_1 and Z_2 are independent and identically distributed given Z . Then,*

$$\mathbb{E}^2 [g_1(Z_1)g_2(Z_2)] \leq \mathbb{E} [g_1(Z_1)g_1(Z_2)] \mathbb{E} [g_2(Z_1)g_2(Z_2)],$$

for any g_1 and g_2 for which the expectations exist, with equality if (but not necessarily only if) g_1 and g_2 are equal up to a multiplicative constant.

Proof. Denote for $i = 1, 2$,

$$\begin{aligned} h_i(z) &= \mathbb{E} [g_i(Z_i) | Z = z] \\ &= \mathbb{E} [g_i(Z_2) | Z = z] \end{aligned}$$

since Z_1 and Z_2 are i.i.d. given Z . Then,

$$\begin{aligned} \mathbb{E} [g_1(Z_1)g_2(Z_2)] &= \mathbb{E} [\mathbb{E} [g_1(Z_1)g_2(Z_2) | Z]] \\ &= \mathbb{E} [h_1(Z)h_2(Z)]. \end{aligned}$$

Following the same steps, we also have,

$$\mathbb{E} [g_i(Z_1)g_i(Z_2)] = \mathbb{E} [h_i^2(Z)], \quad i = 1, 2.$$

The claim now follows by applying the Cauchy-Schwarz inequality to obtain

$$\mathbb{E}^2 [h_1(Z)h_2(Z)] \leq \mathbb{E} [h_1^2(Z)] \mathbb{E} [h_2^2(Z)].$$

□

Example 1. *Let Z be a vector of length N whose entries are i.i.d. Bernoulli(p) and similarly, let W_1 and W_2 be two independent random vectors (of length N) whose entries are i.i.d. Bernoulli(q). Finally, let $Z_i = Z \oplus W_i$ for $i = 1, 2$, where \oplus denotes the binary exclusive or operation. Then the lemma holds for any two functions $g_i : \{0, 1\}^N \rightarrow \mathbb{R}$.*

Remark 2. *We note the method that is used in [4] to prove the maximal correlation property of bivariate normal random variables utilizes series expansions involving (probabilists') Hermite polynomials via Mehler's formula, similar to the approach taken in Lemma 1. In contrast, the proof in Lemma 2 follows the approach taken in [9], where an alternative proof to the maximal correlation property is derived.*

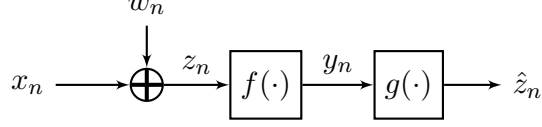


Fig. 1: Memoryless non-linearity operating on a signal corrupted by AWGN.

IV. APPLICATION: A CRITERION FOR IDENTIFICATION OF A MEMORYLESS NON-LINEARITY

We now consider an application of Lemma 1. Consider a memoryless non-linearity f operating on a discrete-time signal corrupted by additive white Gaussian noise (AWGN), as depicted in Figure 1. Thus, the input z_n consists of the sum of the signal x_n and AWGN w_n having variance σ_w^2 . The output is thus,

$$y_n = f(x_n + w_n).$$

We assume that we observe both the input sequence x_n as well as the output y_n . The function f on the other hand is unknown and we wish to estimate it.

Let us consider first the case where it is known that f is invertible. A possible means to identify f is as follows. Apply another function g to the output to obtain

$$\begin{aligned} \hat{z}_n &= g(y_n) \\ &= g(f(z_n)) \\ &= h(z_n), \end{aligned}$$

where $h = g \circ f$ is the composition of the functions f and g . Assume now that x_n is an AWGN process as well (i.e., a training sequence drawn according to such statistics) with variance σ_x^2 . We may further assume for simplicity that $\mathbb{E}[h(z_n)] = 0$. Define

$$K_1 = \frac{\mathbb{E}^2[h(z_n)x_n]}{\mathbb{E}[h^2(z_n)]\mathbb{E}[x_n^2]}.$$

Then, since for bivariate normal random variables nonlinear functions cannot increase (the absolute value of) correlation [4], it follows that K_1 is maximized (only) when $g = c \cdot f^{-1}$ for some constant c , so that h is a linear function. As K_1 may be estimated by replacing expectations with time averages, we have obtained a simple criterion for identification of the non-linearity f (up to to a scale factor that may easily be subsequently estimated).

A limitation of the identification criterion described above, is that it does not apply to non-linearities that are not invertible.¹ We now outline how the inequality derived in this note may serve to overcome this limitation. We note, however, that a drawback of the system described next is that we need to assume that the signal-to-noise ratio σ_x^2/σ_w^2 at the input of the non-linearity is known, unlike for the system described above.

Let $\alpha = \sqrt{\frac{\sigma_x^2 + \sigma_w^2}{\sigma_x^2}}$ so that αx_n has the same variance as z_n . Consider now passing αx_n through a non-linearity g as depicted in Figure 2, to obtain $u_n = g(\alpha x_n)$. Define

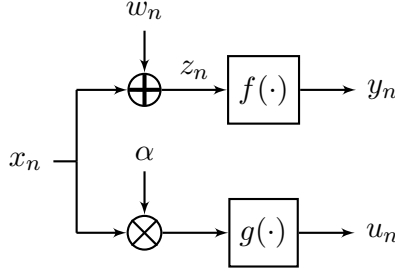


Fig. 2: A system for identification of a non-linearity.

$$K_2 = \frac{\mathbb{E}^2[f(z_n)g(\alpha x_n)]}{\mathbb{E}[g(z_n)g(\alpha x_n)]}.$$

It follows from Lemma 1 that K_2 is maximized *only when* $g = c \cdot f$ for some constant c .² Again, K_2 may be computed by replacing expectations with time averages. This is the case, as although we do not observe w_n , we can replace it with AWGN noise w'_n generated with the same variance, to compute

$$\mathbb{E}[g(z_n)g(\alpha x_n)] = \mathbb{E}[g(x_n + w'_n)g(\alpha x_n)].$$

We note that for a practical implementation of the scheme, one would need some explicit (parametric) representation for the function $g(\cdot)$. For instance, one could employ a series expansion in orthogonal polynomials (see, e.g., [10]).

ACKNOWLEDGMENT

The authors are grateful to Or Ordentlich for helpful discussions concerning the contents of Section IV.

¹Nonetheless, it's a suitable criterion for memoryless nonlinear compensation, where the inverse of the non-linearity (if exists) is desired (see, e.g., [10]).

²Note that the “only when” property, which is crucial for the identification problem, follows by the condition for equality in Lemma 1.

REFERENCES

- [1] J. J. Bussgang, Cross-correlation function of amplitude-distorted Gaussian signals, Tech. rep., Research Lab. of Electronics, M.I.T. (1952).
- [2] R. Price, A useful theorem for nonlinear devices having Gaussian inputs, *Information Theory, IRE Transactions on* 4 (2) (1958) 69–72.
- [3] W. McGee, Circularly complex Gaussian noise—a Price theorem and a Mehler expansion (corresp.), *Information Theory, IEEE Transactions on* 15 (2) (1969) 317–319.
- [4] H. O. Lancaster, Some properties of the bivariate normal distribution considered in the form of a contingency table, *Biometrika* (1957) 289–292.
- [5] H. Rowe, Memoryless nonlinearities with Gaussian inputs: Elementary results, *Bell System Technical Journal* 61 (7) (1982) 1519–1526.
- [6] F. G. Mehler, Ueber die Entwicklung einer Function von beliebig vielen Variablen nach Laplaceschen Functionen höherer Ordnung., *Journal für die reine und angewandte Mathematik* 66 (1866) 161–176.
- [7] W. Kibble, An extension of a theorem of Mehler’s on Hermite polynomials, in: *Mathematical Proceedings of the Cambridge Philosophical Society*, Vol. 41, 1945, pp. 12–15.
- [8] G. Szegő, *Orthogonal Polynomials*, Vol. 23, American Mathematical Soc., 1939.
- [9] Y. Yu, On the maximal correlation coefficient, *Statistics & Probability Letters* 78 (9) (2008) 1072–1075.
- [10] J. Tsimbinos, K. V. Lever, Nonlinear system compensation based on orthogonal polynomial inverses, *Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on* 48 (4) (2001) 406–417.